

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ



Zentrum für
Qualitätssicherung
und -entwicklung

Handreichung der Prüfungswerkstatt

Klausur

André Beem, M.A.

LOB LEHREN – ORGANISIEREN – BERATEN



Bundesministerium
für Bildung
und Forschung



INHALT DIESER HANDREICHUNG

Steckbrief Klausur	2
Definition Klausur	3
Effekte auf Kompetenzförderung und -erfassung	3
Planung und Durchführung von Klausuren	5
Best Practice: Deep-Learning fördern trotz MC-Fragen	9
Checkliste Klausur	11
Literaturverzeichnis	12

STECKBRIEF KLAUSUR

Prüf- bzw. förderbare Kompetenzen
Fachkompetenzen, Methodenkompetenzen

Taxonomiestufe
Wissen erinnern, verstehen, anwenden, Transfer und Beurteilen von Inhalten

Potenziale
Prüfbarkeit fachlicher und methodischer Fähigkeiten, Transfer und Praxisbezug von Lehrinhalten

Herausforderungen
Gefahr der Förderung von Surface-Learning statt Deep-Learning

Gruppengröße
Kleine, mittlere und große Gruppen

Vorbereitungsarbeiten
Prüfungsstoff hinsichtlich der Lernziele analysieren, Beurteilungskriterien festlegen und kommunizieren, Fragen/Aufgaben vorbereiten, Aufsicht und Raum organisieren

Durchführungsarbeiten
Klare Handlungsanweisung für die Bearbeitung der Prüfungsaufgaben geben, Prüfungsaufsicht, ggf. Nachfragen für die gesamte Gruppe beantworten

Nachbereitungsarbeiten
Archivierung, Klausureinsicht, Feedback, Prüfungsdurchlauf reflektieren und evtl. anpassen

DEFINITION KLAUSUR

Eine Klausur zeichnet sich in der Regel durch folgende Merkmale aus: die Prüflinge werden in Gruppen (a) zeitgleich (b) innerhalb eines vorgegebenen Zeitrahmens (c) unter Aufsicht (d) geprüft. Da die Prüflinge zeitgleich und selbstständig die jeweiligen Prüfungsaufgaben bearbeiten, ist dieses Prüfungsformat je nach Verfügbarkeit der Arbeitsplätze auch für große Kohorten geeignet (Schaper/Hilkenmeier 2013; Universität Zürich: AfH 2007; Dysthe et al. 2007; Scouller 1998; Walzik 2012).

Generell lassen sich zwei Untertypen von Klausuren definieren: zum einen Klausuren, bei denen geschlossene Fragestellungen verwendet werden, und zum anderen solche, die eine oder mehrere offene Fragestellungen beinhalten. Bei erstgenanntem Typ müssen die Prüflinge je Frage aus mehreren vorgegebenen Antwortmöglichkeiten auswählen. Daher wird die Prüfungsform auch als Antwortwahlverfahren oder Multiple-Choice bezeichnet. In der zweiten Variante hingegen müssen die Prüflinge eine Schreibleistung erbringen, die in der Regel in dem Verfassen eines Textes, aber auch im Anfertigen einer grafischen oder tabellarischen Darstellung bestehen kann (Universität Zürich: AfH 2007). Häufig wird in der Praxis eine Kombination aus offenen und geschlossenen Fragen in schriftlichen Prüfungen angewendet.

Eine Sonderform bildet die E-Klausur. Hierbei werden die Aufgaben nicht mit Stift und Papier, sondern an einem Computer bearbeitet. Das Einbinden von Video-, Audio- oder Bilddateien, der geringere Korrekturaufwand sowie der vereinfachte Zugang zu den Ergebnissen stellen die wesentlichen Unterschiede zu Paper & Pencil-Klausuren dar. Diese Prüfungsform eignet sich u. a. aufgrund ihrer automatisierten Bewertung und der damit verbundenen Zeitersparnis bei der Korrektur insbesondere für Prüfungen mit hohen Teilnehmerzahlen (Lindner et al. 2015; Wannemacher 2007).

Der Schwerpunkt dieser Handreichung liegt auf Klausuren mit Fragetypen im geschlossenen und offenen (Kurz-) Format. Die Gestaltung und didaktischen Effekte ausführlicherer Essay- bzw. Schreibaufgaben werden in einer gesonderten Handreichung behandelt

EFFEKTE AUF KOMPETENZFÖRDERUNG UND -ERFASSUNG

Nachfolgend werden die Effekte einer Klausurprüfung auf die Erfassung und Förderung verschiedener Kompetenzen beschrieben. Hierbei muss jedoch beachtet werden, dass diese Effekte abhängig vom Formattyp, dem Einsatzszenario und der Aufgabengestaltung sind. Im Folgenden werden sowohl allgemeine Annahmen als auch empirische Evidenzen zu Effekten von Klausurprüfungen vorgestellt.

Fachkompetenzen

Schriftliche Prüfungen eignen sich sowohl mit offenen als auch geschlossenen Frageformaten zur Erfassung von Fachkompetenzen. Insbesondere die Fähigkeit der Reproduktion theoretischen und methodischen Faktenwissens lässt sich in diesem Prüfungsformat ermitteln. Ist die Aufgabenstellung entsprechend formuliert, zum Beispiel durch das Einbinden realer Fallbeispiele, können zum einen ein hoher Praxisbezug hergestellt und auch Kompetenzen auf höheren Taxonomieniveaus (z. B. anwenden und beurteilen) erfasst werden (Hauer 2011; Universität Zürich: AfH 2007; Schaper/Hilkenmeier 2013).

Methodenkompetenzen

Sofern die Prüfungsfragen im offenen Format gehalten sind und somit das eigenständige Verfassen eines Textes erfordern, können auch Methodenkompetenzen, wie wissenschaftliches Schreiben, gefördert und erfasst werden. Dazu zählt u. a. die Verwendung fachspezifischen/wissenschaftlichen Vokabulars und die Reduktion umfangreicher Texte auf die Kernaussagen (Hauer 2011). MC-Aufgaben hingegen lassen aufgrund der vorgegebenen Antwortmöglichkeiten keine Erfassung

von kreativen und schöpferischen Leistungen wie bspw. das Schreiben einer Geschichte zu (Lindner et al. 2015).

Zudem können Klausuren, die überwiegend aus MC-Fragen bestehen, Surface-Learning begünstigen. Demnach zeigen Studierende bei der Vorbereitung auf reine MC-Tests eher oberflächliches als vertiefendes Lernverhalten und schneiden damit erfolgreicher ab als Studierende, die sich im Modus des Deep-Learning vorbereiten (Dysthe et al. 2007; Scouller 1998). Die Prüflinge konzentrieren sich auf die reine Reproduktion der Lerninhalte. Analyse- und Transferkompetenzen, wie die Fähigkeit, Informationen aus verschiedenen Quellen miteinander in Bezug zu setzen und vorhandenes Wissen auf neue Fragestellungen zu übertragen, kann durch Surface-Learning behindert werden (Brauns/Schubert 2008).

Häufig wird in diesem Zusammenhang vom sogenannten Bulimielernen gesprochen. Zurückzuführen ist dieses Phänomen auf die extrinsische Motivation von Prüflingen, sich ausschließlich auf das Bestehen der abschließenden Prüfung zu fokussieren. Insbesondere wenn nur einzelne Lehrinhalte geprüft werden, kann eine Haltung entstehen, in der es im Wesentlichen darum geht, Teilbereiche abzuhaken und dann zu vergessen (Walzik 2012).

SURFACE-LEARNING FÜHRT ZU BESSEREN ERGEBNISSEN BEI MC-KLAUSUREN

Karen Scouller (1998) hat in ihrer Studie den Einfluss der Prüfungsform auf das Lernverhalten von Studierenden untersucht. Dabei stellte sie die Ergebnisse von Prüfungen im offenen Essay-Format solchen im geschlossenen MC-Format gegenüber und brachte diese mit dem gezeigten Lernverhalten (Deep vs. Surface-Learning) in Verbindung.

Untersuchungsdesign

Das der Studie zugrundeliegende Sample bestand aus 206 Lehramtsstudenten/-innen im zweiten Studienjahr, die zwar dieselbe Lehrveranstaltung besucht hatten aber auf unterschiedliche Weise geprüft wurden (Essay-Aufgabe vs. MC-Klausur). Im ersten Schritt wurden die Studierenden hinsichtlich ihrer a) gewählten Lernstrategien und Motive sowie des b) erwarteten Anspruchsniveaus der Prüfungsform und ihrer c) bevorzugten Prüfungsmethode in Bezug auf die Lehrveranstaltungsinhalte befragt. Der Einfluss dieser Variablen auf das Ergebnis im jeweiligen Prüfungsformat wurde anschließend im Rahmen einer Regressionsanalyse ermittelt.

Zentrale Ergebnisse

- Studierende, die sich auf eine MC-Klausur vorbereiten, wählen eher Surface-Learning Strategien. Während Studierende, die sich auf eine Prüfung im Essay-Format vorbereiten, eher Deep-Learning Strategien wählen.
- Studierende gehen davon aus, dass mit MC-Klausuren eher niedrigere kognitive Level, wie die Reproduktion von Wissen geprüft werden, während bei Prüfungen im Essay-Format erwartet wird, dass höhere kognitive Level wie das Analysieren komplexerer Sachverhalte geprüft werden.
- Studierende, die bei der Vorbereitung auf eine MC-Klausur Deep-Learning Strategien wählen, schneiden schlechter ab als solche Studierende, die Surface-Learning Strategien anwenden.
- Studierende, die bei der Vorbereitung auf eine Prüfung im Essay-Format Surface-Learning Strategien wählen, schneiden schlechter ab als solche Studierende, die Deep-Learning Strategien anwenden (S. 461-463).

PLANUNG UND DURCHFÜHRUNG VON KLAUSUREN

Um den studentischen Lernerfolg bestmöglich zu fördern, sollten Prüfungsform und -inhalte mit den Lernzielen sowie den Lehr-Lernaktivitäten einer Lehrveranstaltung abgestimmt werden („constructive alignment“; siehe Handreichung „Kompetenzorientiertes Prüfen“). Entsprechend sollten die folgenden Aspekte der Planung und Konzeption einer Prüfung bereits vor Veranstaltungsbeginn im Rahmen der Veranstaltungsplanung mitbedacht werden.

a. Allgemeine Prüfungsvorbereitung (vor Veranstaltungsbeginn)

• Ökonomische Aspekte abwägen

Generell gilt, dass der Zeit- und Arbeitsaufwand für die Planung und Durchführung schriftlicher Prüfungen i. d. R. vergleichsweise gering ist (z. B. im Gegensatz zu mündlichen Prüfungen, vgl. Huxham et al. 2012). Ein Großteil des Zeit- und Arbeitsaufwandes muss in die Gestaltung der Prüfungsfragen und der Bewertungskriterien investiert werden. Diese sollte idealerweise im (Experten-)Team stattfinden. Bei E-Klausuren ist zudem zu beachten, dass neben der inhaltlichen Gestaltung der Aufgaben die technische Umsetzung zusätzlichen Arbeitsaufwand verursacht. Je nach Kenntnisstand und Fähigkeiten im Umgang mit der entsprechenden E-Prüfungsplattform (z. B. ILIAS) muss genügend Zeit für Training und Beratung der Klausurautoren/-innen einkalkuliert werden. Der Arbeitsaufwand für die Bewertung von geschlossenen Aufgaben ist i.d.R. gering, da diese standardisiert und bei E-Klausuren sogar vollkommen automatisiert durchgeführt werden kann. Die Bewertung von offenen Aufgaben hingegen ist zeitintensiver und aufwendiger.

• Klausurtyp wählen

In der Klausurgestaltung kann wie beschrieben generell zwischen offenen und geschlossenen Fragetypen unterschieden werden. Geschlossene Fragetypen eignen sich besonders für das Prüfen von Fach- und Methodenkompetenzen auf den unteren Taxonomiestufen, während mit offenen Fragetypen auch höhere Taxonomiestufen und kreative Schreibkompetenzen geprüft werden können. Durch entsprechende inhaltliche Gestaltung der einzelnen Fragen und/oder eine Kombination verschiedener Fragetypen lässt sich dieses Paradigma jedoch auflockern (Hauer 2011; Universität Zürich: AfH 2007; Schaper/Hilkenmeier 2013). Eine weitere Form der schriftlichen Prüfung ist die E-Klausur, die sowohl offene als auch geschlossene Fragetypen beinhalten kann und deren Einsatz bspw. dann sinnvoll sein kann, wenn Multimedia-Dateien (z. B. Audioaufnahmen, Röntgenbilder) Bestandteil der Prüfung sein sollen.

• Vor der Klausur Übungsklausuren anbieten

Als sinnvoll haben sich Trainingsphasen in Form von Probeklausuren erwiesen, die der eigentlichen Prüfung vorgeschaltet sind. Schwierigkeit und Aufgabenstruktur sollte dabei in etwa der tatsächlichen Abschlussklausur entsprechen. Bei E-Klausuren können solche vorbereitenden Angebote zudem verhindern, dass sich individuell stark abweichende Computerkenntnisse negativ auf den Prüfungsablauf und die Prüfungsergebnisse auswirken (Wannemacher 2007).

b. Prüfungskonzeption und Aufgabengestaltung (vor Veranstaltungsbeginn)

Generell gilt, dass sowohl die Prüfungskonzeption als auch die konkrete Aufgabengestaltung zum Zweck der Qualitätssicherung im (Experten-)Team durchgeführt werden sollte. Der gesamte Prozess sollte auch

nach durchgeführter Prüfung nicht als abgeschlossen angesehen werden und stetig im Rahmen kollegialen Austauschs fortgeführt werden. Dies fördert nicht zuletzt die Entwicklung gemeinsamer Standards zur Leistungsmessung in einem Studiengang.

- **Hohe Objektivität/Reliabilität erreichen**

Hinsichtlich der Objektivität von Prüfungen wird zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden. Um eine höchstmögliche Objektivität zu erreichen, sollte jeder Prüfungsteilnehmer unter den gleichen Testbedingungen geprüft werden, die Auswertung möglichst standardisiert erfolgen und das Ergebnis der jeweiligen Prüfung anhand einheitlicher Kriterien bewertet werden. Diese Bedingungen lassen sich speziell bei Prüfungen mit vornehmlich geschlossenen Fragen besonders erfüllen (Brauns/Schubert 2008; Downing/Haladyna 2004).

- **Rateeffekte vermindern**

Bei MC-Fragen kann die Ratewahrscheinlichkeit einen negativen Einfluss auf die Reliabilität und als Folge dessen auch auf die Validität der einzelnen Prüfungsfragen haben. Die Effektstärke der Ratewahrscheinlichkeit kann durch eine Erhöhung der Anzahl von Testfragen und Antwortmöglichkeiten abgemindert werden (Zimmermann/Williams 2003). Auch die sorgfältige Konstruktion der MC-Fragen (eindeutiger Arbeitsauftrag, Trennschärfe) und der dazugehörigen Antwortmöglichkeiten kann zu einer Verminderung dieses Effekts führen (Downing/Haladyna 2004).

- **Gestaltung der Distraktoren**

Die Auswahl geeigneter Distraktoren (Falschantworten) ist die wichtigste Voraussetzung für eine praktikable MC-Aufgabe und eine faktische Begrenzung der Ratewahrscheinlichkeit. Ein Distraktor ist geeignet, wenn er Personen ohne erforderliches Wissen plausibel erscheint, von

Personen mit den erforderlichen Fähigkeiten und Kenntnissen jedoch als falsch erkannt wird. Er ist ungeeignet, wenn er auch ohne erwartetes Wissen leicht als Distraktor erkannt und ausgeschlossen werden kann (Lindner et al. 2015). Häufig sind der dritte und vierte Distraktor so offensichtlich falsch, dass sie auch von den leistungsschwächsten Prüfungsteilnehmern nicht ausgewählt werden (Brauns/Schubert 2008). MC-Fragen mit einer korrekten und zwei falschen Antwortmöglichkeiten sind in Bezug auf Schwierigkeitsgrad, Diskrimination sowie Reliabilität und Trennschärfe am effektivsten (Downing/Haladyna 2004; Rodriguez 2005; Lee et al. 2011).

- **Steuerung der Performanzmessung in den Randbereichen**

MC-Aufgaben differenzieren insbesondere in den Randbereichen der Performanzmessung schlechter als offene Aufgabentypen zwischen ansonsten leistungsstarken und -schwachen Prüfungsteilnehmern/innen. Prüflinge im höheren Leistungsbereich schneiden bei offenen Aufgabenformaten besser ab als solche im niedrigeren Leistungsbereich. Demnach werden Prüflinge im unteren Leistungsbereich durch die Möglichkeit, richtige Antworten bei MC-Fragen zu erraten, eher zu vorteilhaft bewertet, während leistungsstarke Studierende ihr Potenzial nicht vollständig zeigen können (Lee et al. 2011; Lindner et al. 2015).

- **Schwierigkeitsgrad variieren**

Allgemein gilt, dass die Schwierigkeit offener Items die Schwierigkeit geschlossener Items übersteigt (Berger/Bouley 2015). Um Deckeneffekte zu vermeiden bzw. einen möglichst effektiv zwischen leistungsstarken und -schwachen Studierenden differenzierenden Test zu erstellen, sollten die Schwierigkeiten der einzelnen Testfragen variieren (Zeuch 2011).

- **Ermüdungseffekte reduzieren**

Beim Lernen geht es um ein zunehmend konsistentes Antwortverhalten sowie um das Vermögen, mehrere Dimensionen gleichzeitig in ein Urteil zu integrieren. Ermüdungs- und Langeweile-Effekte schlagen sich in einer sinkenden Konsistenz und in einer Fokussierung auf weniger Merkmale oder andere vereinfachte Entscheidungsregeln nieder. Eine Aneinanderreihung von konzeptionell ähnlichen Fragen (z. B. gleicher Fragetyp, gleiche Schwierigkeit) kann diesen Effekt verstärken und sollte vermieden werden (Auspurg et al. 2009).

- **Didaktisch sinnvoller Medien- und Softwareeinsatz in E-Klausuren**

Die barrierefreie Möglichkeit der nahtlosen Einbindung von Bild-, Ton- und Videoquellen in computergestützten Prüfungen ist ein wesentlicher Unterschied zu konventionellen Klausuren. Hierbei ist jedoch darauf hinzuweisen, dass der Einsatz solcher Medienobjekte immer dem didaktischen Ziel unterzuordnen ist und nicht zum „Aufhübschen“ der Fragen verwendet werden sollte (Hauer 2011; Lindner et al. 2015). Zudem sollten Lehrinhalte, die den praktischen Einsatz von Computerprogrammen (z. B. SPSS, Excel) beinhalten, auch am Computer geprüft werden, um einen Medienbruch zwischen Lehre und Prüfung zu vermeiden (Cella et al. 2007; Steiner 2009; Pueschel 1999; Wanemacher 2007).

c. Durchführung (während der Lehrveranstaltung bzw. Prüfung)

- **Didaktisches Ziel und Anforderungen der Prüfung kommunizieren**

Um den Prüflingen Handlungssicherheit zu geben und den erwünschten Lernprozess zu fördern, sollten die didaktischen Überlegungen, die die Wahl der Prüfungsform begründen, sowie Arbeitsauftrag und Bewertungskriterien eindeutig

definiert und den Studierenden rechtzeitig vor der Klausur, idealerweise bereits zu Beginn der Lehrveranstaltung, kommuniziert werden. Dabei sollten auch die Aufgabenstellung sowie die erlaubten Hilfsmittel bzw. zugrundeliegende Informationen unmissverständlich formuliert sein und die maximal erreichbaren Punkte öffentlich gemacht werden.

- **Rechtliche Rahmenbedingungen einhalten**

Unmittelbar vor der Prüfung sind Identität, Prüfungszulassung und Prüfungsfähigkeit der Studierenden abzufragen.

- **Verhalten bei Täuschungsversuchen**

Sowohl dem Aufsichtspersonal als auch den Prüflingen sollten vor Klausurstart klare Handlungsanweisungen im Fall von Täuschungsversuchen gegeben werden.

- **Feedback geben**

Um den Kompetenzerwerb zu fördern und den studentischen Lernprozess optimal zu begleiten, ist strukturiertes Feedback unverzichtbar. In dieser Hinsicht ist eine reine Rückmeldung in Form einer Abschlussnote nach einer Klausurprüfung unzulänglich. Im Idealfall sollte zu mehreren Zeitpunkten des Lernprozesses ein kriteriengestütztes (individuelles) Feedback erfolgen. In Veranstaltungen, die mit einer summativen schriftlichen Prüfung abgeschlossen werden, sind nachfolgende Varianten denkbar:

- **Veranstaltungsbegleitendes Feedback (z.B. mittels Probeklausur) anbieten**

Bereits vor der Prüfung sollten die Studierenden Feedback erhalten, um ihren individuellen Lernfortschritt und ihre Lernmethoden reflektieren und ggf. Wissenslücken schließen zu können. Formative anonyme Tests oder unbenotete Probeklausuren können hier Ausgangspunkt für eine individuelle Rückmeldung sein.

- **Schriftliches Feedback geben**

Das Feedback kann schriftlich in Form von kurzen einzelnen Kommentaren und Erläuterungen am Antworttext oder auch als Zusammenfassung gegeben werden. Ein konkreter Bezug auf die Antwort und die Darlegung der Bewertungskriterien ist in beiden Fällen essentiell.

- **Mündliches Feedback geben**

Individuelles mündliches Feedback kann bspw. im Rahmen der Klausureinsicht erfolgen und ermöglicht eine direkte Interaktion mit den Studierenden. Eventuelle Unklarheiten oder Nachfragen bezüglich der Bewertung können so unmittelbar thematisiert werden.

- **Automatisiertes Feedback in E-Klausuren konfigurieren**

In E-Klausuren kann ein automatisiertes Feedback konfiguriert werden. So kann bspw. bei einer falsch beantworteten MC-Frage eine kurze Erläuterung mit ergänzenden Informationen zum Inhalt der Falschantwort angezeigt werden.

d. Bewertung

• **Bewertungsfehler vermeiden**

Die Bewertung offener Fragen sollte generell kriterienbasiert sein. Hierfür sollte im Rahmen der Prüfungsvorbereitung ein Kriterienraster erstellt werden. Auf diese Weise können Bewertungsobjektivität gesteigert und Beurteilungsfehler wie beispielsweise Halo- und Reihenfolgeeffekte vermindert werden. Für das Erstellen eines solchen Kriterienrasters ist zusätzliche Vorbereitungszeit einzuplanen. Bei handschriftlichen Klausuren kommt der Faktor der möglichen Unleserlichkeit hinzu.

• **Transparenz schaffen**

Die Bewertung und Punktevergabe sollte für die Prüflinge eindeutig nachvollziehbar sein.

e. Nachbereitung

• **Klausureinsicht anbieten**

Den Prüfungsteilnehmern/innen muss die Möglichkeit der Klausureinsicht gestattet werden. Bei E-Klausuren kann eine Einsicht nach vorheriger Raumbuchung in den PC-Pools des ZDV durchgeführt werden. Generell sollten keine Musterlösungen preisgegeben werden, da die entsprechenden Fragen dann für eine etwaige wiederholte Verwendung unbrauchbar werden.

• **Archivierung sicherstellen**

Sowohl Prüfungsfragen als auch die Prüfungsergebnisse müssen archiviert werden. Die Archivierung von E-Klausuren mit ILIAS wird vom E-Learning-Team übernommen. Sollten nachträglich Anpassungen bei Bewertungen oder am Notenschema vorgenommen werden, muss das E-Learning-Team informiert werden, damit stets die aktuelle Klausurversion gespeichert werden kann.

• **Evaluation durchführen**

Nach der Klausur sollte zum Zweck der Qualitätssicherung eine Evaluation der Prüfungsfragen durchgeführt werden. Eine Itemanalyse liefert beispielsweise statistische Werte über Messgenauigkeit, Trennschärfe und Schwierigkeit einzelner Prüfungsfragen.

BEST PRACTICE: DEEP-LEARNING FÖRDERN TROTZ MC-FRAGEN

Die Kritik, dass mit MC-Fragen ausschließlich die Reproduktion auswendig gelernter Fakten geprüft werden kann und somit Surface-Learning begünstigt wird, ist unberechtigt. Welche Kompetenzen und Taxonomiestufen abgefragt werden, hängt von der Aufgabengestaltung und der Kreativität der Aufgabenautoren ab. Zudem können Fragen, die auf Wissensreproduktion abzielen, auch durchaus einen didaktischen Nutzen (z. B. als Eisbrecherfrage) erfüllen. Dass allerdings auch im geschlossenen Fragenformat höhere Taxonomiestufen und verschiedene Kompetenzen geprüft werden können, sollen die nachfolgenden Praxisbeispiele verdeutlichen.

Beispiel 1:

Intendierte Kompetenz: Fachspezifische theoretische Kenntnisse

Intendierte Taxonomiestufe: Erinnern

What is the average effective radiation dose from chest CT? (1P)

- a. 1 mSv
- b. 8 mSv
- c. 16 mSv
- d. 24 mSv

Best Practice weil...

...durch die Angabe der max. punktzahl Bewertungstransparenz hergestellt wird.

...die Studierenden erlerntes fachspezifisches Wissen erinnern müssen und somit exakt die intendierte Kompetenz auf der angestrebten Taxonomiestufe abgeprüft wird.

Beispiel 2:

Intendierte Kompetenz: Fachspezifische theoretische Kenntnisse

Intendierte Taxonomiestufe: Anwenden

A 30-year-old man presented with a 4-month history of dyspnea, low-grade fever, cough, and fatigue. Given the following chest radiograph, what is the most likely diagnosis? (1P)

- a. Adult respiratory distress syndrome
- b. Pulmonary edema
- c. Streptococcal pneumonia
- d. Pulmonary alveolar proteinosis

Best Practice weil...

...direkter Berufs-/Praxisbezug in der Vignette.

...die Antwortmöglichkeiten eindeutig voneinander abgrenzbar sind.

Beispiel 3:

Intendierte Kompetenz: Transferfähigkeit

Intendierte Taxonomiestufe: Beurteilen

Read the following comments a teacher made about testing. Then answer the question that follows by circling the letter of the best answer.

“Students go to school to learn, not to take tests. In addition, tests cannot be used to indicate a student’s absolute level of learning. All tests can do is rank students in order of achievement, and this relative ranking is influenced by guessing, bluffing, and the subjective opinions of the teacher doing the scoring. The teacher-learning process would benefit if we did away with tests and depended on student self-evaluation.”

Which one of the following propositions is most essential to the final conclusion?
(1P)

- a. Effective self-evaluation does not require the use of tests.
- b. Tests place students in rank order only.
- c. Test scores are influenced by factors other than achievement.
- d. Students do not go to school to take tests.

Best Practice weil...

...konkrete Anweisungen für Handlungssicherheit sorgen.

...die Studierenden die Aussage in einen theoretischen Kontext einfügen müssen und somit nicht nur Wissen reproduzieren.

CHECKLISTE KLAUSUR

a. Allgemeine Prüfungsvorbereitung

- Lassen sich mit dem Prüfungsformat die in den Lernzielen enthaltenen studentischen Handlungsfähigkeiten erfassen?
- Wurden die (prüfungs-)rechtlichen Vorgaben eingehalten?
- Ist der Aufwand für Vorbereitung und Durchführung ökonomisch sinnvoll?
- Wurde eine Übungsklausur angeboten?

b. Prüfungskonzeption und Aufgabengestaltung

- Wurden Maßnahmen zur Erhöhung der Objektivität und Reliabilität ergriffen?
- Wurden bei der Entscheidung über die Anzahl der Prüfungsfragen/Antwortmöglichkeiten eventuelle Rateeffekte miteinbezogen?
- Wurde der Schwerpunkt bei der Gestaltung der Distraktoren auf deren Qualität und nicht auf die Quantität gelegt?
- Wurden verschiedene Fragentypen eingesetzt, um Ermüdungseffekte zu minimieren?
- Unterscheiden sich die Prüfungsfragen hinsichtlich des Schwierigkeitsgrads?
- Wurde bei der Wahl der Aufgabengestaltung Aspekte der Performanzmessung bzw. die Stärken und Schwächen der einzelnen Fragentypen in dieser Hinsicht beachtet?
- Ist der Einsatz von Medien und Software in E-Klausuren didaktisch sinnvoll?

c. Durchführung

- Wurden die Handlungsanweisungen bezüglich der Aufgabenbearbeitung unmissverständlich formuliert?
- Sind sowohl den Prüflingen als auch den Aufsichtspersonen die Konsequenzen bei Täuschungsversuchen bewusst?
- Wurde die Identität der Klausurteilnehmer überprüft?
- Wurde der Lernprozess zu mehreren Zeitpunkten mit Feedbackelementen angeleitet?
- Wurde das Prüfungsergebnis mittels eines schriftlichen oder mündlichen Feedbacks erläutert?

d. Bewertung

- Ist die Bewertung kriterienbasiert erfolgt?
- Besteht Transparenz hinsichtlich der Bewertungskriterien?

e. Nachbereitung

- Wurde eine Klausureinsicht angeboten?
- Wurden sowohl Testfragen als auch -ergebnisse archiviert?
- Wurden die Prüfungsfragen zum Zweck der Qualitätssicherung statistisch ausgewertet?

LITERATURVERZEICHNIS

- Alderson, J. Charles; Wall, Dianne (1993): Does washback exist? In: *Applied Linguistics* (14 (2)), S. 115–129, zuletzt geprüft am 20.03.2017.
- Auspurg, Katrin; Hinz, Thomas; Liebig, Stefan (2009): Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey 2009, zuletzt geprüft am 02.03.2017.
- Berger, Stefanie; Bouley, Franziska (2015): Fachwissen und fachdidaktisches Wissen im wirtschaftspädagogischen Studium - Entwicklung eines Testinstruments und erste empirische Befunde. In: Barbara Koch-Priewe, Anne Köker, Jürgen Seifried und Eveline Wuttke (Hg.): *Kompetenzerwerb an Hochschulen. Modellierung und Messung zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*. Bad Heilbrunn: Verlag Julius Klinkhardt, S. 105–128.
- Brauns, Katrin; Schubert, Sebastian (2008): Qualitätssicherung von Multiple-Choice-Prüfungen. In: Sigrid Dany (Hg.): *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen*. Bielefeld: Bertelsmann (Blickpunkt Hochschuldidaktik, 118), S. 92–102.
- Cella, David; Gershon, Richard; Lai, Jin-Shei; Choi, Seung (2007): The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. In: *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 16 Suppl 1, S. 133–141.
- Collins, Jannette (2006): Education techniques for lifelong learning. Writing multiple-choice questions for continuing medical education activities and self-assessment modules. In: *Radiographics : a review publication of the Radiological Society of North America, Inc* 26 (2), S. 543–551. DOI: 10.1148/rg.262055145.
- Downing, Steven M.; Haladyna, Thomas M. (2004): Validity threats. Overcoming interference with proposed interpretations of assessment data. In: *Med Educ* 38 (3), S. 327–333.
- Dysthe, Olga; Engelsen, Knut Steinar; Lima, Ivar (2007): Variations in portfolio assessment in higher education. Discussion of quality issues based on a Norwegian survey across institutions and disciplines. In: *Assessing Writing* 12 (2), S. 129–148.
- Harlen, Wynne & Deakin Crick, Ruth (2003): "Testing and Motivation for Learning". In: *Assessment in Education*, 10 (2).
- Hauer, Erich (2011): Wird dumm geprüft, wird dumm gelernt. Plädoyer für den Einsatz anwendungsorientierter Prüfungsaufgaben im Hochschulbereich. In: *Magazin Erwachsenenbildung.at* (2011) 12, 10 S. Online verfügbar unter http://www.pedocs.de/volltexte/2013/7418/pdf/Erwachsenenbildung_12_2011_Hauer_Wird_dumm_geprueft.pdf.
- Huxham, Max; Campbell, Fiona; Westwood, Jenny (2012): "Oral versus written assessments: a test of student performance and attitudes". In: *Assessment and Evaluation in Higher Education* 37 (1), 125-136.
- Lee, Hee-Sun; Liu, Ou Lydia; Linn, Marcia C. (2011): Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. In: *Applied Measurement in Education* 24 (2), S. 115–136. DOI: 10.1080/08957347.2011.554604.
- Lindner, Marlit A.; Strobel, Benjamin; Köller, Olaf (2015): Multiple-Choice-Prüfungen an Hochschulen? In: *Zeitschrift für Pädagogische Psychologie* 29 (3-4), S. 133–149.
- Pueschel, Constantin (1999): *Potentiale von Web-Videokonferenzsystemen bei akademischen Prü.* Universität Konstanz. Konstanz.
- Rodriguez, Michael C. (2005): Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. In: *Educational Measurement issues and practice*, 24, S. 3–13.
- Schaper, Niclas; Hilkenmeier, Frederic (2013): *Umsetzungshilfen für kompetenzorientiertes Prüfen*. HRK-Zusatzgutachten
- Sambell, Kay; McDowell, Lit & Brown, Sally (1997): „„But is it fair?": an Exploratory Study of Student Perceptions of the Consequential Validity of Assessment." In: *Studies in Higher Education* 23 (4), 349-371.
- Scouller, Karen (1998): The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. In: *Higher Education* 35 (4), S. 453–472.
- Spratt, Mary (2005): „Washback and the classroom: the implications for teaching and learning of studies of washback from exams". In: *Language Teaching Research* 9 (1), 5-29.
- Steiner, Heinke (2009): *Online-Assessment. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis ; mit 42 Tabellen*. 1. Aufl. Heidelberg: Springer Medizin-Verl.
- Universität Zürich; Arbeitsstelle für Hochschuldidaktik AfH, 2007: *Leistungsnachweise in modularisierten Studiengängen*. Universität Zürich, Zürich. Arbeitsstelle für Hochschuldidaktik.
- Walzik, Sebastian (2012): *Kompetenzorientiert prüfen [Elektronische Ressource]*. Leistungsbewertung an der Hochschule in Theorie und Praxis. 1. Aufl. Opladen: Barbara Budrich (UTB, 3777). Online verfügbar unter <http://www.utb-studi-e-book.de/9783838537771>.

- Wannemacher, Klaus (2007): Computerbasierte Prüfungen. Zwischen Self-Assessment und Abschlussklausuren. In: Michael H. Breitner, Beate Bruns und Franz Lehner: Neue Trends im E-Learning [Elektronische Ressource]. Aspekte der Betriebswirtschaftslehre und Informatik. Hg. v. Michael H. Breitner. Heidelberg: Physica-Verlag HD, S. 427–440.
- Zeuch, Nina (2011): Rule based item construction. Analysis with and comparison of linear logistic test models and cognitive diagnostic models with two item types. Online verfügbar unter https://repositorium.uni-muenster.de/document/miami/566ab0b8-465e-4fd0-80d0-a490aae29111/diss_zeuch.pdf, zuletzt geprüft am 01.03.2017.
- Zimmerman, Donald W.; Williams, Richard H. (2003): A New Look at the Influence of Guessing on the Reliability of Multiple-Choice Tests. In: Applied Psychological Measurement 27 (5), S. 357–371.